An Error Estimate for Matrix Equations

Yang Cao and Linda Petzold *

August 26, 2002

Department of Computer Science, University of California Santa Barbara
CA 93106, USA.

Abstract.

This paper proposes a new method for estimating the error in the solution of matrix equations. The estimate is based on the adjoint method in combination with small sample statistical theory. It can be implemented simply and is inexpensive to compute. Numerical examples are presented which illustrate the power and effectiveness of the new estimate.
Keywords: condition number, adjoint method, Lyapunov equation, Sylvester equation.
Mathematics Subject Classification: 65F35, 65F30, 15A12.

1 Introduction

Matrix equations play an important role in applied mathematics and control theory. Examples of well-known matrix equations include the Sylvester equation

$$AX - XB = C, \tag{1}$$

where $A \in R^{n\times n}$, $B \in R^{m\times m}$, $C \in R^{n\times m}$, and $X \in R^{n\times m}$ is to be determined. The Sylvester equation arises in many applications. For example, the finite difference discretization of a separable elliptic boundary value problem on a rectangular domain can be written in the form of a Sylvester equation [20]. The Sylvester equation arises also in the design of reduced order observers [2], and in many eigenproblems [1], [11]. When $B = -A^T$, the Sylvester equation reduces to the Lyapunov equation

$$AX + XA^T = C. \tag{2}$$

Important applications of the Lyapunov equation include model reduction [3], [17], stability analysis of linear systems [18] and in the solution of another important matrix equation: the algebraic Riccati equation

$$A^T X + XA - XGX = W, \tag{3}$$

which is widely applied in optimal control theory [15], [16].

In this paper we are concerned with estimating the error in the solution of matrix equations. First we must specify the norm with which to measure the error. In the literature of perturbation

*This work was supported by grants: NSF/ITR ACI-0086061, NSF/KDI ATM-9873133 and DOE DE-F603-00ER25430.

theory for matrix equations, the Frobenius norm is most often used because it is the vector 2-norm when the system is formulated as a linear system (see [13] as an example). Perturbation theory based on other norms has also been discussed for some special cases. In particular, a highly effective simple method for estimating the error in the spectral norm for the stable Lyapunov equation and the algebraic Riccati equation has been proposed in [9], [12], [16]. We briefly describe this method for the stable Lyapunov equation. Suppose A is stable. Thus all the eigenvalues of A are in the left-half plane. For the Lyapunov equation (2), the solution X has the form

$$X = \int_0^\infty e^{At} C e^{A^T t} dt.$$

Consider the perturbed matrix equation

$$(A + \Delta A)(X + \Delta X) + (X + \Delta X)(A + \Delta A)^T = C + \Delta C.$$

The corresponding error ΔX also has the form

$$\Delta X = \int_0^\infty e^{At}(\Delta C + \Delta A(X + \Delta X) + (X + \Delta X)\Delta A^T) e^{A^T t} dt. \tag{4}$$

For the spectral norm, let u and v denote the left and right singular vectors of unit length of ΔX, such that

$$u^* \Delta X v = \|\Delta X\|.$$

We have

$$\|\Delta X\| \le \int_0^\infty u e^{At}(\Delta C + \Delta A(X + \Delta X) + (X + \Delta X)\Delta A^T) e^{A^T t} v dt.$$

Then the spectral norm condition number is given by [12]

$$K = 2\|A\|\|H\|, \tag{5}$$

where H solves the equation $AH + HA^T = -I$, and $\int_0^\infty \|e^{At}u\|^2 dt \le \|H\|$ for any unit length u. The relative error is bounded by

$$\frac{\|\Delta X\|}{\|X + \Delta X\|} \le K \left[\frac{\|\Delta A\|}{\|A + \Delta A\|} + \frac{\|\Delta C\|}{\|C + \Delta C\|} \right]. \tag{6}$$

The relation (6) is also used in [9] to estimate the error in the Frobenius norm although no proof is given. This error bound is specific to the stable Lyapunov equation, because only with stability do we have the formula (4). For the unstable Lyapunov equation, a counterexample can be found in [9]. There is a formula similar to (4) for the stable Sylvester equation. However, a similar generalization based on that formula requires the solution of the matrix equations $AH_1 + H_1 A^T = -I$ and $BH_2 + H_2 B^T = -I$, which is much costlier than the case of the stable Lyapunov equation.

In this paper we are concerned with the Frobenius norm of the general matrix equation. To simplify the presentation, we will take the Sylvester equation as our example. Numerical results will be given for both the Sylvester and Lyapunov equations.

For the Sylvester equation, it is well known that when A and B have no eigenvalues in common, the Sylvester equation (1) is nonsingular. Numerical algorithms for the solution of the Sylvester equation can be found in [10], [11], [13]. Perturbation theory for the Sylvester equation has been studied as a generalization of perturbation theory for linear systems [8], [13]. For this purpose, (1) is written in the form

$$(I_n \otimes A - B^T \otimes I_m)\mathrm{vec}(X) = \mathrm{vec}(C), \tag{7}$$

where $A \otimes B = (a_{ij}B)$ is the Kronecker product, and the *vec* operator stacks the columns of a matrix into one long vector. An error bound is then formed using the condition number of the linear system (7). Here we have to be very careful because the transformation from (1) to (7) may turn an ill-conditioned system to a well conditioned one. For example, in [8] the condition number is defined as

$$k(A,B) = \frac{\max_i \sigma_i(A \otimes I - I \otimes B^T)}{\min_i \sigma_i(A \otimes I - I \otimes B^T)}. \tag{8}$$

This is the condition number of the linear system (7) but may not give an accurate measurement of the condition of the Sylvester equation (1). Consider the following example

Example 1 Let

$$A = \begin{bmatrix} 1+\epsilon & 0 \\ 0 & 1-\epsilon \end{bmatrix}, \quad B = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \quad C = \begin{bmatrix} \epsilon & 0 \\ 0 & -\epsilon \end{bmatrix},$$

where ϵ is very small. The estimate (8) yields a moderate condition number but the original system (1) is ill-conditioned. The eigenvalues of A are very close to the eigenvalues of B, which makes this system very close to a singular one. The reason for this discrepancy is the poor condition of the transformation from (1) to (7).

The problem of estimating the condition of a matrix equation is considered also in [13]. Consider the perturbed Sylvester equation

$$(A+\Delta A)(X+\Delta X) - (X+\Delta X)(B+\Delta B) = C + \Delta C, \tag{9}$$

where $\|\Delta A\|_F \leq \epsilon\alpha$, $\|\Delta B\|_F \leq \epsilon\beta$, and $\|\Delta C\|_F \leq \epsilon\gamma$. Let $P = I_m \otimes A - B^T \otimes I_n$. After dropping the second-order terms, we obtain

$$A\Delta X - \Delta X B = \Delta C - \Delta A X + X \Delta B.$$

Writing this system in the form

$$P\mathrm{vec}(\Delta X) = -\begin{bmatrix} X^T \otimes I_n & -I_m \otimes X & I_{nm} \end{bmatrix} \begin{bmatrix} \mathrm{vec}(\Delta A) \\ \mathrm{vec}(\Delta B) \\ \mathrm{vec}(\Delta C) \end{bmatrix},$$

the condition number is given by ([13], p. 318)

$$\Phi = \|P^{-1}[\alpha(X^T \otimes I_n) \quad -\beta(I_m \otimes X) \quad -\gamma I_{nm}]\|_2 / \|X\|_F.$$

This can be simplified and weakened to the more often quoted condition number (see [12], for example)

$$\Psi = \|P^{-1}\|_2 \frac{(\alpha+\beta)\|X\|_F + \gamma}{\|X\|_F}.$$

The error bound is then given by

$$\frac{\|\Delta X\|_F}{\|X\|_F} \leq \sqrt{3}\Phi\epsilon, \tag{10}$$

or

$$\frac{\|\Delta X\|_F}{\|X\|_F} \leq \sqrt{3}\Psi\epsilon. \tag{11}$$

It is well known that (10) generates a sharper condition number than (11) but the computation of (10) is more expensive. Both Ψ and Φ yield large condition numbers for example 1, which reflects the ill-conditioning of the system. However, the computational cost to determine $\|P^{-1}\|$ is high. Iterative methods are generally used to estimate this term. Moreover, this definition is based on norm-wise analysis. It assumes that the norm of the perturbations is bounded by ϵ times the norm of the data. This assumption cannot take into account the structure of the perturbation. Thus for structured problems, norm-wise estimates yield an overestimate. For example, consider the following well-conditioned system

Example 2

$$\begin{bmatrix} 2 & 0 \\ 0 & \delta \end{bmatrix} X - X \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & \delta \end{bmatrix}, \tag{12}$$

which has the unique solution $X = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$. The norm-wise condition number gives $\Phi = O(\frac{1}{\delta})$, $\Psi = O(\frac{1}{\delta})$. When δ is small, this is a large overestimate of the error. In this case, we need component-wise analysis, which has been proposed for linear systems [5], [13], [19]. Component-wise analysis scales the perturbation with the data and usually gives a sharper estimate. But it has not yet been applied to the perturbation theory of matrix equations, to the best of our knowledge.

Yet another motivation for an improved error estimation method is the subspace condition. For some problems, we may be concerned with a subspace of the solution rather than the full space. This requires a well-conditioned subspace solution only. In this case, even if the whole system is ill-conditioned, we may still be able to safely obtain a solution in a well-conditioned subspace. Such examples for linear systems can be found in [5], [6].

In this paper we will present a new method for error estimation of matrix equations. It yields the component-wise error estimate and can be used to obtain a subspace error estimate. The computational cost is low, and with high probability the result is accurate to within a factor of 10, which is the usual requirement for error estimates. Numerical experiments show that this method gives a sharper error estimate than (11) for the general Sylvester equation. It is much cheaper than (10), while the accuracy is similar. It yields a much sharper error estimate for matrix equations with some special structures. For the stable Lyapunov equation, our method yields an error estimate which is sharper than (6), but at a higher computational cost. This paper is organized as follows. In Section 2 we introduce the method of adjoint sensitivity analysis for matrix equations. In Section 3 we present the error estimation technique, which is based on adjoint sensitivity analysis and the small sample statistical method[14]. Numerical results are given in Section 4.

2 Sensitivity Analysis for Matrix Equations

We derive our method for the general nonlinear matrix equation. Consider the perturbations as parameters of the system. Then the matrix equation is formulated as

$$F(X, p) = 0, \tag{13}$$

where $X \in R^{m \times n}$, $F : R^{m \times n} \to R^{m \times n}$, and p is a vector which represents the perturbations. The question we are concerned with is how the solution of (13) will change when the corresponding parameters p change. Thus we wish to estimate the magnitude of the corresponding sensitivity X_p.

When there are many components of X and parameters p, the corresponding sensitivity X_p has a high dimension. In this section we show how to use the adjoint method to efficiently compute

the sensitivity of a scalar function $g(X)$ with respect to the parameters p. In the following section, we will show how to use this method together with small sample statistical theory to estimate the magnitude of X_p and hence the error of the solution.

Consider a scalar function $g : R^{m \times n} \to R$. For any parameter p, we have

$$\frac{dg}{dp} = \sum_{i,j} \frac{dg}{dx_{ij}} \frac{dx_{ij}}{dp}. \tag{14}$$

$\frac{dg}{dx_{ij}}$ can be formulated as a matrix, denoted by $\frac{dg}{dX}$. For example, if $g(X) = \|X\|_F$, $\frac{dg}{dX} = \frac{X}{\|X\|_F}$. We will make use of the inner product of two matrices, defined by

$$< X, Y >= tr(X^T Y) = \sum_{i,j} x_{ij} y_{ij}. \tag{15}$$

Note that $\|X\|_F = (< X, X >)^{\frac{1}{2}}$. We can express the sensitivity of $g(X)$ with respect to p as

$$\frac{dg}{dp} =< \frac{dg}{dX}, X_p > . \tag{16}$$

On the other hand, from

$$F(X, p) = 0,$$

we have

$$F_p + \frac{dF}{dX}(X_p) = 0, \tag{17}$$

where $\frac{dF}{dX}$ is the differential operator defined as

$$\frac{dF}{dX}(Y) = \lim_{h \to 0} \frac{F(X + hY) - F(X)}{h}. \tag{18}$$

Taking the inner product of both sides of (17) with a matrix Λ yields

$$< \Lambda, F_p + \frac{dF}{dX}(X_p) >= 0,$$

thus

$$< \Lambda, \frac{dF}{dX}(X_p) >= - < \Lambda, F_p > .$$

Now the basic idea of the adjoint method is to choose Λ so that

$$< \Lambda, \frac{dF}{dX}(X_p) >=< \frac{dg}{dX}, X_p > . \tag{19}$$

Then the sensitivity can be computed easily by

$$\frac{dg}{dp} = - < \Lambda, F_p > . \tag{20}$$

To accomplish this, let the adjoint operator $(\frac{dF}{dX})^*$ of $\frac{dF}{dX}$ be defined by

$$< Z, \frac{dF}{dX}(Y) >=< (\frac{dF}{dX})^*(Z), Y > . \tag{21}$$

Now solve the adjoint equation

$$(\frac{dF}{dX})^*(\Lambda) = \frac{dg}{dX} \tag{22}$$

for the matrix Λ. Then the sensitivity is given by (20).

For the Sylvester equation, we have $F(X) = AX - XB - W$. The differential operator is given by

$$\frac{dF}{dX}(Y) = AY - YB.$$

Thus

$$< Z, \frac{dF}{dX}(Y) >=< Z, AY - YB >=< A^T Z - ZB^T, Y > .$$

The adjoint operator is given by

$$(\frac{dF}{dX})^*(Z) = A^T Z - ZB^T,$$

and the adjoint equation is

$$A^T \Lambda - \Lambda B^T = \frac{dg}{dX}. \tag{23}$$

The sensitivity $\frac{dg}{dp}$ is then given by $< \Lambda, A_p X - X B_p - W_p >$.

3 Error Estimate

We use sensitivity analysis to obtain the error estimate. The perturbations to the data are considered as parameters of the system. The error estimate is based directly on the sensitivity to these perturbations. Note that we do not need to be able to compute the sensitivity very accurately. We seek to estimate the magnitude of the error to within a factor of 10.

The perturbations arise from data error or from round-off error. Either relative or absolute error may be of the most concern, depending on the situation. Relative error is usually of the greatest concern in numerical solution. Thus in the following discussion we will focus on the relative error. The absolute error estimate is easy to obtain in a similar manner.

First let us consider the change in a scalar function g with respect to the perturbations. We have

$$|g(X(0)) - g(X(p))| \approx \sum_p |\frac{dg}{dp}||p|$$

when the perturbations $|p|$ are small. Suppose $|p| < \epsilon$. Then

$$|g(X(0)) - g(X(p))| \leq \sum_p |\frac{dg}{dp}|\epsilon, \tag{24}$$

where $\frac{dg}{dp}$ is given by (20). Note that Λ is independent of p, and

$$\sum_p |\frac{dg}{dp}| = \sum_p | < \Lambda, F_p > |. \tag{25}$$

Then (24) and (25) give the error estimate for the scalar function g.

Our objective is to estimate the error in the solution of a matrix equation. The error is a multi-dimensional function of the solution. However, to use a multi-dimensional function to compute the estimate would be too expensive. Thus a strategy for making use of scalar functions to estimate

the error of the vector (matrix) is needed. For the general case, we make use of the small sample statistical method introduced in [14].

Consider ΔX as a vector. One way to estimate $\|\Delta X\|_F$ is to take its inner product with a vector randomly selected from the unit ball S_{k-1}, where k is the dimension of ΔX. For the Sylvester equation, $k = nm$. Let the random matrix selected from the unit sphere S_{k-1} be R. According to [14], we have

$$E(| < R, \Delta X > |) = E_k \|\Delta X\|_F,$$

where E denotes the expectation, and $E_1 = 1$, $E_2 = \frac{2}{\pi}$, and for $k > 2$,

$$E_n = \frac{1 \cdot 3 \cdot 5 \cdots (k-2)}{2 \cdot 4 \cdot 6 \cdots (k-1)} \quad \text{for } k \text{ odd},$$

$$E_n = \frac{2}{\pi} \cdot \frac{2 \cdot 4 \cdot 6 \cdots (k-2)}{1 \cdot 3 \cdot 5 \cdots (k-1)} \quad \text{for } k \text{ even}.$$

E_k can be estimated by $\sqrt{\frac{2}{\pi(k-\frac{1}{2})}}$. $\|\Delta X\|_F$ can be estimated, using one random matrix, by

$$\|\Delta X\|_F \approx \frac{| < R, \Delta X > |}{E_k}.$$

Thus we can define the scalar function $g(X) =< R, X >$. Then $\Delta g =< R, \Delta X >$. Since g is a scalar function, the sensitivity estimate from Section 2 can be applied.

The corresponding probability for one random matrix is given by

$$Pr\left(\frac{1}{w}\|\Delta X\|_F \leq \frac{| < R, \Delta X > |}{E_k} \leq w\|\Delta X\|_F\right) \approx 1 - \frac{2}{\pi w}.$$

Suppose we require an error estimate which is accurate to within a factor of 10. Then for one random matrix, the probability of a good estimate is about 93%. If we need a higher probability or a more accurate estimate, more random matrices can be used. Usually we use 2 or 3 random matrices. Let R_i be orthogonal random matrices selected from S_{k-1}. $< R_i, R_j >= 0$ if $i \neq j$. Define $v_i = | < R_i, \Delta X > |$. The error estimates are given by

$$e_2 = \frac{E_2\sqrt{v_1^2 + v_2^2}}{E_k}$$

with 2 random matrices and

$$e_3 = \frac{E_3\sqrt{v_1^2 + v_2^2 + v_3^2}}{E_k}$$

with 3 random matrices. The corresponding probabilities are

$$Pr\left(\frac{1}{w}\|\Delta X\|_F \leq e_2 \leq w\|\Delta X\|_F\right) \approx 1 - \frac{\pi}{4w^2}$$

and

$$Pr\left(\frac{1}{w}\|\Delta X\|_F \leq e_3 \leq w\|\Delta X\|_F\right) \approx 1 - \frac{32}{3\pi^2 w^3}.$$

Letting $w = 10$, using 2 random matrices gives probability 99.21%, while using 3 random matrices gives probability 99.89%. $v = | < R, \Delta X > |$ is estimated by solving the adjoint equation (22). Then (24) and (25) yield the bound. Let Λ be the solution of (22) and $u = \sum_p | < \Lambda, \frac{dF}{dp} > |$. Then $|v| \leq |u|\epsilon$. Note that the computational cost is not high because we can use the matrix decomposition which was formed when solving the original system. For example, the Schur decomposition used in the solution of the Sylvester equation can be used for solving the corresponding adjoint equation. Below we describe the algorithm using 2 random matrices.

Error estimate algorithm with 2 random matrices Suppose we have solved the original matrix equation (1) and the corresponding matrix decomposition is available.

Step 1: Randomly choose 2 orthogonal matrices R_1, R_2 from the unit sphere S_{k-1}. Solve (22) for the corresponding Λ_1, Λ_2.

Step 2: Compute

$$u_i = \sum_p | < \Lambda_i, \frac{dF}{dp} > |.$$

Then the absolute error estimate for $\|\Delta X\|_F$ is given by

$$e = \frac{E_2\sqrt{(u_1^2 + u_2^2)}}{E_k}\epsilon. \tag{26}$$

The relative error estimate for $\frac{\|\Delta X\|_F}{\|X\|_F}$ is given by $\frac{e}{\|X\|_F}$.

By choosing the scalar function in an appropriate manner, we can generalize the above algorithm to produce a subspace error estimate. Suppose we want to estimate the error of $Y = PXQ$, where $P \in R^{p\times n}$, $Q \in R^{m\times q}$ with $p << n, q << m$. We first select a random matrix $R \in R^{p\times q}$ uniformly from the unit sphere S_{k-1}, where $k = p \times q$. Then, defining $g(X) =< R, PXQ >$, the adjoint method and small sample statistical method can be applied as before. Note that in this case $\frac{dg}{dX} = P^T R Q^T$.

Sylvester equation For the Sylvester equation, the sensitivity with respect to perturbations in C is easy to compute. It is bounded by $< |\Lambda|, |C| >$. For perturbations in A, considering the relative error, let $\delta a_{ij} = \epsilon_{ij} a_{ij}$. Then

$$\frac{\partial F}{\partial \epsilon_{ij}} = A_p X = a_{ji} e_{ij} X,$$

and

$$< \Lambda, \frac{\partial F}{\partial \epsilon_{ij}} >= \sum_k [a_{ji}\lambda_{ik} x_{jk}].$$

The sensitivity with respect to perturbations in B can be obtained similarly. Thus

$$\begin{array}{rcl} \sum_p |\frac{dg}{dp}| & = & < |\Lambda|, |W| > + \sum_{ij} |\sum_k a_{ji}\lambda_{ik} x_{jk}| + \sum_{ij} |\sum_k b_{kj}\lambda_{kj} x_{ki}| \\ & \leq & < |\Lambda|, |A||X| + |X||B| + |C| > . \end{array} \tag{27}$$

Remark: We should comment on the difference between this method and the original small sample statistical estimation for matrix equations as in [14]? The method in [14] generates the random vectors from the perturbation space, while our method generates the random vectors from the solution space. For example, for a linear system $Ax = b$, [14] generates random vectors Z_A, Z_b with unit length (dimension $n^2 + n$), then multiplies them by a small number δ and solves the system $(A + \delta Z_A)(x + \Delta x) = b + \delta z_b$. Thus $\frac{\|\Delta x\|}{E_{n^2+n}\delta}$ is the condition number estimated. The choice of δ is critical and there is no good rule for it. Our method generates random vectors z with unit length (dimension n) from the solution space, constructs a derived function $g(x) = z^T x$, and then solves the adjoint equation. Thus we avoid the critical choice of δ, and we generate lower dimension random vectors.

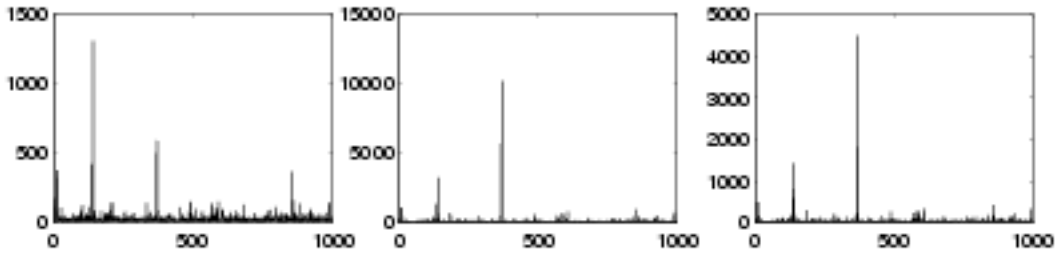

Figure 1: Overestimate ratio of our method $\frac{our\ error\ estimate}{actual\ relative\ error}$ (left), condition estimate (11) $\frac{error\ estimate\ (11)}{actual\ relative\ error}$ (middle), and condition estimate (10) $\frac{error\ estimate\ (10)}{actual\ relative\ error}$ (right) for 1,000 Sylvester equations with randomly generated dense matrices A, B and X of dimension 10.

4 Numerical Experiments

The numerical experiments were performed in Matlab on a Linux computer. We chopped the data so that the round-off error is 10^{-8}, in order to avoid any possibility of differences in the conclusions that could be caused by different machine precisions.

The numerical experiments are based on randomly generated data. We compare our error estimate with the condition error estimate (10), (11), and, when applicable, with the method (6) proposed in [12] for the stable Lyapunov equation. We first generate the random matrices A, B and X. Then C is determined by $C = AX - XB$. We chop the data of A, B and C to a relative error of 10^{-8}. Then we solve $A\tilde{X} - \tilde{X}B = C$. We compare the error estimates and the actual relative error $\frac{\|X-\tilde{X}\|_F}{\|X\|_F}$. For the error estimate (10) and (11), $\|P^{-1}\|$ is computed accurately without approximation.

When we consider the error estimate, every possible perturbation is taken into account. This usually yields an overestimate of the relative error. In all of our experiments, seldom is an underestimate generated. Thus we only compare the overestimate ratio $\frac{\text{estimate}}{\text{actual relative error}}$.

1. Dense Sylvester Equation In Figure 1, we compare the overestimate ratio of different estimates for 1000 randomly generated matrices A, B and X of dimension 10. The mean value of the overestimate ratio of our method is 30.73, while the mean value of the overestimate ratio of (10) is 43.12 and the mean value of the overestimate ratio of (11) is 98.28. In the extreme case, (10) gives an overestimate ratio as high as 4,486, and (11) gives an overestimate ratio as high as 10,200, while the highest overestimate ratio given by our method is 1,308. We can see that generally our method gives a sharper estimate. As we have mentioned, the computational cost for our method is also lower.

2. Diagonal Matrices If the Sylvester equation has some special structure, our method can give much better results. In this experiment we randomly generate diagonal matrices A and B. Because of the special structure, the actual relative error is small. The overestimate ratio is plotted in Figure 2. The mean value of the overestimate ratio of our method is 11.58, while the mean value of the overestimate ratio of (10) is 17.96, the mean value of the overestimate ratio of (11) is 60.37. In the extreme case, (11) gives an overestimate ratio as high as 2,467, and (10) gives an overestimate ratio as high as 722, while the highest overestimate ratio given by our method is 19.44.

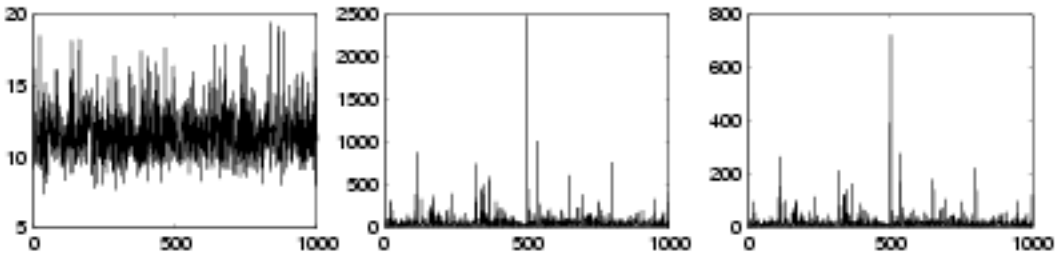

Figure 2: Overestimate ratio of our method $\frac{our\ error\ estimate}{actual\ relative\ error}$ (left), condition estimate (11) $\frac{error\ estimate\ (11)}{actual\ relative\ error}$ (middle), and condition estimate (10) $\frac{error\ estimate\ (10)}{actual\ relative\ error}$ (right) for 1,000 Sylvester equations with randomly generated diagonal matrices A and B, and dense matrices X of dimension 10.

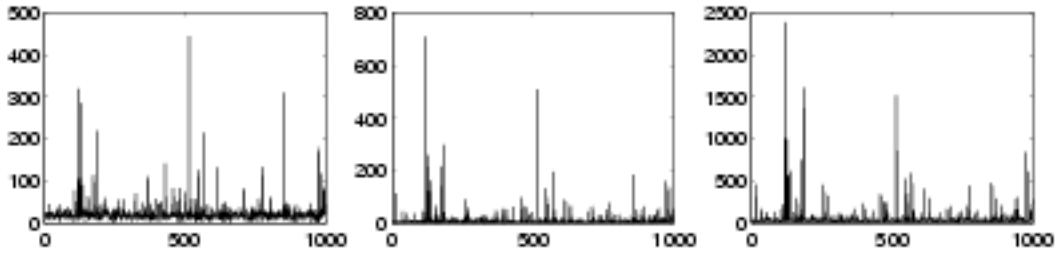

Figure 3: Overestimate ratio of our method (left), condition estimate (10) (middle) and the method (6) of [12] (right) for 1,000 Lyapunov equations with randomly generated stable matrices A, and dense matrices X of dimension 10.

3. Stable Lyapunov Equation For the stable Lyapunov equation, [12] gives a simple error estimate (6). In this experiment, we randomly generate 1,000 stable matrices A of dimension 10. The overestimate ratio of (6) is compared with our method and the condition estimate method (10) (since (10) is sharper than (11)). The result is plotted in Figure 3. The mean value of the overestimate ratio of our method is 22.90, while the mean value of the overestimate ratio of (10) is 12.31, and the mean value of the overestimate ratio of (6) is 53.00. We can see that our method gives a sharper estimate than (6) but we note that the computational cost of (6) is lower unless we use only one random matrix to generate our error estimate method (but then it is less reliable).

5 Conclusion

In this paper we proposed a new method for error estimation for matrix equations, based on the adjoint sensitivity method and small-sample statistical theory. Our method has low computational cost, and probability of 99% for the accuracy of the error estimate to be within a factor of 10. Numerical experiments show that our method yields sharper estimates for randomly generated dense Sylvester equations than the standard error estimate. For stable Lyapunov equations, our method's computational cost is higher than that of the error estimate proposed in [12], but the estimate is sharper. Our method can be generalized in a straightforward manner to other matrix equations, for example, the discrete-time Lyapunov equation and the algebraic Riccati equation.

Acknowledgments The authors would like to thank Professor Karl Åstrom and Charles Kenney for their insightful comments in discussions related to this work.

References

[1] Z. Bai and J. Demmel, *On swapping diagonal blocks in real Schur form*, Lin. Alg. Appl., Vol. 186, 1993, 73-96.

[2] J. Barlow, M. Monahemi and D. O'Leary, *Constrained matrix Sylvester equations*, SIAM J. Matrix Anal. Appl., Vol. 13, 1992, 1-9.

[3] D. Bernstein and D. Hyland, *The optimal projection equations for reduced-order state estimation*, IEEE Trans. Automat. Contr., Vol. AC-30, 1985, 583-585.

[4] R. Byers, *Numerical condition of the Algebraic Riccati equation*, in Proc. Summer Research Conference, AMS Vol. 47, Contemporary Math., American Mathematical Society, Providence, RI, 1984, 35-49.

[5] Y. Cao and L. Petzold, *A subspace error estimate for linear systems*, to appear, SIAM. J. Matrix Anal. and Appl.

[6] S. Chandrasekaran and I. C. Ipsen, *On the sensitivity of solution components in linear systems of equations*, SIAM J. Matrix Anal. Appl., Vol. 16, 1 (1995), 93-112.

[7] A. K. Cline, C. B. Moler, G. W. Stewart and J. H. Wilkinson, *An estimate for the condition number of a matrix*, SIAM J. Numer. Anal., Vol. 16, 2 (1979), 368-375.

[8] A. Deif, N. Seif and S. Hussein, *Sylvester's equation: accuracy and computational stability*, J. Comput. Appl. Math., Vol. 61, 1995, 1-11.

[9] P. Gahinet, A. Laub, C. Kenney and G. Hewer, *Sensitivity of the stable discrete-time Lyapunov equation*, IEEE Trans. Automat. Control, AC-35(11): 1209-1217, 1990.

[10] J. Gardiner, A. Laub, J. Amato and C. Moler, *Solution of the Sylvester matrix equation* $AXB^T + CXD^T = E$, ACM Trans. Math. Software, 18(2), 1992, 223-231.

[11] G. H. Golub and C. F. Van Loan, *Matrix Computations*, 2nd edition, the Johns Hopkins University Press, 1989.

[12] G. Hewer and C. Kenney, *The sensitivity of the stable Lyapunov equations*, SIAM J. Cont. Opt., Vol. 26, 2, (1988), 321-344.

[13] N. J. Higham, *Accuracy and Stability of Numerical Algorithms*, SIAM, 1996.

[14] C. S. Kenney and A. J. Laub, *Small-sample statistical condition estimates for general matrix functions*, SIAM J. Sci. Comput., Vol. 15, 1, (1994), 36-61.

[15] C. Kenney and G. Hewer, *The sensitivity of the algebraic and differential Riccati equations*, SIAM J. Cont. Opt., Vol. 28, 1, (1990), 50-69.

[16] A. Laub, *A Schur method for solving algebraic Riccati equations*, IEEE Trans. Automat. Control, AC-24(6): 913-921, 1979.

[17] B. Moore, *Principal component analysis in linear systems: Controllability, observability, and model reduction*, IEEE Trans. Automat. Contr., Vol. AC-26, 1981, 17-31.

[18] L. Salle and S. Lefschetz, *Stability by Liapunov's Direct Method*, New York, Academic, 1961.

[19] R. D. Skeel, *Scaling for numerical stability in Gaussian elimination*, JACM. Vol. 26, 3, (1979), 494-526.

[20] G. Starke and W. Niethammer, *SOR for AX - XB = C*, Linear Alg. Appl., Vol. 154, 1991, 355-375.